



Dealing with Diversity in Computational Cancer Modeling

Citation

Johnson, David, Steve McKeever, Georgios Stamatakos, Dimitra Dionysiou, Norbert Graf, Vangelis Sakkalis, Konstantinos Marias, Zhihui Wang, and Thomas S. Deisboeck. 2013. Dealing with diversity in computational cancer modeling. *Cancer Informatics* 12: 115-124.

Published Version

doi:10.4137/CIN.S11583

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:11181088>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

OPEN ACCESS

Full open access to this and thousands of other papers at <http://www.la-press.com>.

Dealing with Diversity in Computational Cancer Modeling

David Johnson¹, Steve McKeever², Georgios Stamatakis³, Dimitra Dionysiou³, Norbert Graf⁴, Vangelis Sakkalis⁵, Konstantinos Marias⁵, Zhihui Wang⁶ and Thomas S. Deisboeck⁷

¹Department of Computer Science, University of Oxford, Oxford, UK. ²Department of Informatics and Media, Uppsala University, Uppsala, Sweden. ³Institute of Communication and Computer Systems, National Technical University of Athens, Athens, Greece. ⁴Department of Paediatric Haematology and Oncology, Saarland University Hospital, Homburg, Germany. ⁵Institute of Computer Science at the Foundation for Research and Technology—Hellas, Heraklion, Crete, Greece. ⁶Department of Pathology, University of New Mexico, Albuquerque, NM, USA. ⁷Harvard-MIT (HST), Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, MA, USA.
Corresponding author email: david.johnson@cs.ox.ac.uk

Abstract: This paper discusses the need for interconnecting computational cancer models from different sources and scales within clinically relevant scenarios to increase the accuracy of the models and speed up their clinical adaptation, validation, and eventual translation. We briefly review current interoperability efforts drawing upon our experiences with the development of in silico models for predictive oncology within a number of European Commission Virtual Physiological Human initiative projects on cancer. A clinically relevant scenario, addressing brain tumor modeling that illustrates the need for coupling models from different sources and levels of complexity, is described. General approaches to enabling interoperability using XML-based markup languages for biological modeling are reviewed, concluding with a discussion on efforts towards developing cancer-specific XML markup to couple multiple component models for predictive in silico oncology.

Keywords: multi-scale computational tumor modeling; in silico oncology; model interoperability; XML markup languages

Cancer Informatics 2013;12 115–124

doi: [10.4137/CIN.S11583](https://doi.org/10.4137/CIN.S11583)

This article is available from <http://www.la-press.com>.

© the author(s), publisher and licensee Libertas Academica Ltd.

This is an open access article published under the Creative Commons CC-BY-NC 3.0 license.



Introduction

The last few decades have witnessed an increased interest of the scientific community in the development of computational models for simulating tumor growth and response to treatment.^{1–8} The major modeling techniques can be differentiated as predominantly continuous and predominantly discrete models. Continuous models rely primarily on differential equations to describe processes such as diffusion of molecules, changes in tumor cell density, and invasion of tumor cells into the surrounding tissue.^{9–14} Discrete modeling considers several discrete states in which cells may be found and possible transitions between them governed by “decision calculators” such as cytokinetic diagrams and agent-based techniques.^{15–20} Discrete models are usually represented by cellular automata of several forms and variable complexity. Due to the hypercomplexity²¹ of cancer-related topics, each modeling approach is intrinsically able to satisfactorily address only some of the aspects of this multifaceted problem.

In recent years, data-driven computational cancer modeling has become an active field in cancer research.²² In particular, the development of cancer models that encompass different biological scales in time and space (ie, multiscale cancer models) has gained attention in view of the potential to integrate disparate kinds of patient data and to enable patient-specific prediction and assist in treatment planning.^{23–25} Consequently, these techniques fall into two basic types of approaches to cancer modeling: “bottom-up” and “top-down.” The bottom-up approach studies the components of a system individually and then integrates the properties and functions of each component to make predictions about the behavior of the entire system.²⁶ On the other hand, a top-down approach, driven by observed biological characteristics or phenomena, builds up theories that would explain the observed behavior.²⁶ In the particular case of cancer simulation and prediction, agent-based modeling (ABM) has been widely adopted as a useful technique for developing bottom-up models, whereas both discrete and continuum modeling are used for developing top-down models. Combining both techniques yields a “hybrid” approach.

The bottom-up approach is suitable for simulating emergent cancer behaviors resulting from cell-cell and cell-host interactions and intracellular signaling

of individual cells. Many bottom-up multiscale cancer models have been developed so far where most of them are based on the ABM technique and incorporate a specific molecular-level description. Recent representative examples include those quantifying the relationship between extracellular growth factors and multicellular cancer growth and expansion,^{27–30} those investigating cancer cell motility in an evolving tumor population by connecting gene regulatory networks to cell phenotypes,^{31–33} those describing genotype-phenotype relations based on and studying the effects of different cell adhesion pathways on cancer cell invasion patterns.^{34,35} All these models explicitly access and draw on prior knowledge about biochemical and biophysical mechanisms and the underlying biological properties of cancer. This demonstrates the potential of the bottom-up approach in making full use of the sizeable amount of molecular and microscopic data being generated experimentally and in clinics. A top-down simulation approach^{15,36,37} typically starts from the macroscopic imaging data (a high biocomplexity level) and proceeds toward lower biocomplexity levels. When there is a need for an upward movement in the biocomplexity scales, a summary of the available information pertaining to the previous lower level is used. A top-down approach is suitable for directly simulating clinical trials, and therefore clinically adapting, validating, and eventually translating the models into clinical practice. It also offers the possibility to exploit the actual multiscale data of the individual patient, including molecular markers.

To better understand and subsequently treat cancer more effectively, a significant effort is underway to develop and use models of cancer pathophysiology in order to simulate cancer evolution and promote individualized, that is, patient-specific optimization of, disease treatment. The latter is leading to a central clinical question from the context of predictive oncology: Is it possible to select the best targeted therapy for a patient by computer simulation?

To answer this question and be able to promote predictive oncology, it is mandatory to validate cancer models in real clinical cases and assess the added value in optimizing therapy selection for the individual patient in studies. It is important to keep in mind that cancer is a multiscale phenomenon, and while many research groups develop significant models, they

usually address only specific scales (from molecular to tissue level) and are difficult to integrate due to the fact that there is no universally accepted standard for presenting and implementing such models. Another adverse effect of such compartmentalized research is that there is no established protocol for the clinical validation of cancer models or for the assessment of their results. In particular, the lack of standardized descriptions of models significantly hampers their widespread adoption and clinical testing and, more importantly, their interconnection in order to efficiently couple models of different scales and improve their accuracy.

This paper draws on our experiences with the VPH³⁸ (Virtual Physiological Human, <http://www.vph-noe.eu>) projects on cancer modeling: ACGT³⁹ (Advancing Clinico Genomic Trials on Cancer, eu-acgt.org), ContraCancrum⁶ (Clinically Oriented Cancer Multilevel modelling, <http://www.contracancrum.eu>) and the US CViT⁴⁰ (Center for the Development of a Virtual Tumor, <http://www.cvit.org>) projects. We also look to the future with the developments within the TUMOR⁴¹ (Transatlantic Tumor Model Repositories, <http://www.tumour-project.eu>) project that is paving the way for an integrated, interoperable transatlantic research environment and investigating standards for simulation and modeling within the domain of predictive in silico oncology.

Clinically Oriented in Silico Oncology

Sophisticated multiscale models yield valuable quantitative insights into complex mechanisms involved in cancer and may ultimately contribute to patient-specific therapy optimization. The ultimate goal of clinically oriented cancer simulation models is their eventual translation into clinical practice, which entails two key steps. Firstly, thorough sensitivity analyses are carried out in order to both comprehend and validate model behaviors. This will enable researchers to gain further insights into the simulated mechanisms in a more quantitative way. Secondly, an adaptation and validation process based on real clinical data is carried out.

The clinical orientation of a model constitutes a fundamental guiding principle throughout its development. In order to ultimately support clinical decision making in a patient-individualized manner, clinically oriented models should be under continuous

refinement within the framework of clinical trials. For a clinician it is important that the in silico experiments can address and answer precisely for each patient the following questions: What is the natural course of the tumor growth over time in size and shape? When and where to is the tumor metastasizing? Can the response of the local tumor and the metastases to a given treatment be predicted in size and shape over time? What is the best treatment schedule in terms of drugs, surgery, irradiation, and their combination, dosage, time schedule and duration to achieve a positive outcome? Is it possible to predict severe adverse events of a treatment and to propose alternatives to them without jeopardizing the outcome? Is it possible to predict a cancer before it occurs and to recommend treatment options to prevent the occurrence or a recurrence? The question to be addressed would be decided by the clinician and consequently influence the model. Clinically oriented in silico oncology seeks to address such questions.

An example clinical scenario

To exemplify the need for coupling models together for clinically-oriented in silico oncology, in this section we describe a clinical scenario that combines two distinctly different approaches for different purposes to increase the accuracy of a diagnosis.

Malignant gliomas (World Health Organization [WHO] grade III and IV) are progressive brain tumors that can be divided into anaplastic gliomas (WHO grade III) and glioblastoma multiforme (GBM) (WHO grade IV) based upon their histopathologic features.^{42,43} Magnetic resonance imaging (MRI) has become the method of choice in the diagnostic workup of these patients.⁴⁴ Because of edema surrounding the tumor and the presence of necrotic and vital areas within the tumor, the exact tumor volume is nearly impossible to define.⁴⁵ This is true at the time of diagnosis and even much more enhanced during treatment, as edema and necrotic areas might change with an increase in tumor volume despite treatment response (pseudoprogression).

Magnetic resonance spectroscopy, diffusion weighted imaging, as well as perfusion MRI, can depict changes in the cellular metabolism. Positron emission tomography is used to detect tumors with high metabolic rates of glucose. These commonly used imaging modalities still pose problems when



identifying pseudoprogression and pseudoregression in clinical practice. On average, patients who suffer from grade III gliomas have an average survival of 2 to 3 years. In contrast, most patients with GBM die of the disease within a year following diagnosis.^{42,46} Over the last few years the most important improvement has been achieved by the concomitant and adjuvant application of temozolomide and radiotherapy, which increased the survival period from approximately 12 to 15 months.⁴⁷ Long-term survivors for more than 5 years are sparse (3%–5% of GBM patients).⁴⁸ Improvement in outcome of patients with GBM is urgently needed and can only be achieved through a combined effort between clinicians, basic researchers, computer scientists, mathematicians, and legal and ethical policy-makers. New treatment modalities need to be developed.

The simulation of GBM *in silico* is one such option by modeling tumor growth and response to treatment. Since cancer is a strongly multiscale natural phenomenon, in order to be able to provide reliable predictions of its spatiotemporal course, including response to treatment modalities, several biocomplexity scales should be addressed concurrently in a combinatorial way. This implies that sufficiently advanced models of several biomechanisms concerning different spatiotemporal scales have to be developed and adequately coupled. Different modeling groups worldwide focus on different scales and contexts of tumor dynamics. Therefore, the models they produce have in general different external and internal structures. Coupling such models tends to be a highly demanding task. The combination of a bottom-up approach with a top-down approach will combine data from systems biology such as cell cycle duration or methylation status of *MGMT* (a DNA repair gene correlating with outcome after temozolomide treatment⁴⁹) or deregulated metabolic pathways with real patient data such as age, appearance of the tumor in imaging studies, and outcome. As a goal for the future, such spatiotemporal multidimensional models have to be integrated into daily clinical care and need to provide validated results for single patients in due time. Clinicians using such models should be guided by a closed workflow that encapsulates patient data provision, preprocessing and postprocessing of data (including anonymization or pseudonymization of patient-identifiable data), uploading and integration

of data, and computational execution of the chosen *in silico* model. The validated results after model execution are then sent to the clinician to assist in the treatment decision-making process. For example, consider the following possible models of treatment response in GBM that may be calculated computationally: (1) Differentiation between real progression and pseudoprogression after irradiation of the tumor and (2) Simulation of the response to a combined treatment with irradiation and temozolomide.

The correct assessment of response to a given treatment is difficult to assess. An unspecific disruption of the blood-brain barrier may cause a reactive treatment-related edema mimicking tumor progression.⁴⁴ This is often seen after irradiation and summarized as pseudoprogression. On the other hand, pseudoregression is also found if treatment (eg, antiangiogenic drugs) is not affecting the tumor itself but the surrounding edema.⁵⁰ While a predictive simulation of treatment response in the second model provides a means to assisting therapy, the differentiation of real and pseudo responses, as in the first model, could be integrated. These two models coupled together might increase the accuracy of an *in silico* prediction of GBM treatment response.⁵¹

An Overview of Major Interoperability Efforts

The need for coupling models together has been highlighted by the clinical scenario described in the previous section, but how can we facilitate connecting disparate models together? Markup languages for modeling biological systems (based on the Extensible Markup Language [XML]^{52,53}) emerged in the early 2000s to address the problems associated with the lack of standards for describing biological models. Four major languages have gained prominence in recent years each of which aims to tackle the problems associated with interoperability of models. The markup languages discussed here are SBML, CellML, FieldML and *insilicoML*.

Systems biology markup language

The Systems Biology Markup Language, commonly referred to as SBML, is a domain-specific markup language that addresses biochemical processes at the molecular scale.^{54,55} The motivations for

SBML were 3-fold: (1) to support multiple tools with a single common file format, (2) to enable repeatability of experiments with published models irrespective of modeling software platform, and (3) to promote longevity of published models beyond the lifetime of current modeling tools. These aims are quite generalized; however, the authors explain that SBML does not aim to be a generic modeling language to cover all types of quantitative models. They recognize that the *de facto* understanding of different biological concepts evolve, and, as such, they submit that a modeling language for systems biology be domain-specific and is structured to represent the consensus of current understanding in the field. This aims to enable the state-of-the-art modeling tools in systems biology to use a common language in which to communicate models rather than having a single generic modeling language for biological and/or computational modeling.

To describe the mathematical components in SBML, the language utilizes Content MathML, an XML language for describing mathematical formulae.⁵⁶ Typically, the mathematics used to model systems biology is in the form of declarative formulae such as ordinary differential equations (ODEs) and partial differential equations (PDEs), and the markup used can adequately describe such equations. Models are structured as sets of components termed compartments that broadly represent containers for chemical substances. Changes in the values and states within compartments are dictated by description statements of biochemical transformations or transport. It also provides the facility to associate metadata with models in order to properly curate them within online databases. The details of the latest release of SBML (Level 3) are extensively described in Hucka et al.⁵⁵

CellML

Developed out of the physiological modeling community, CellML is a modeling markup language that aims to cover a range of biological phenomenon, primarily cell function.^{57,58} CellML was developed to address the lack of standards for describing cellular function and to provide unambiguous representations of models. The authors identified that because of the lack of rigor and standards in the publishing process, models could not be easily validated. Errors are commonly introduced when publishing models in

journal texts, and computational implementations are commonly targeted at specific software frameworks and tools, making the models themselves less portable. This poses problems when sharing with researchers who are unfamiliar with the modeling methodologies, frameworks, and tools others may have used.

Like SBML, CellML utilizes Content MathML to describe systems modeled using mathematical equations. CellML is designed to be modular in that encapsulated models (possibly of different scales) can be linked together through public and private interfaces. This allows multiple models whose variables might refer to the same entity can be logically linked. This component-based approach allows reuse of whole models or parts of models described with CellML markup. To compliment CellML's functional description of biological cells, FieldML is being developed as a language for modeling physiological structures based on geometric meshes and fields, allowing the representation of spatial variation and PDEs.⁵⁹ Structures are represented as abstractions of physical states over locations described and approximated sets of functions.

insilicoML

insilicoML (ISML) is a markup language for describing biological models developed out of the Japanese Physiome Project.^{60–62} ISML was developed to be a modular description of models and has a number of similarities to CellML. The authors designed the language with a set of tools that facilitate conversion to multiple representational formats such as CellML, SBML, general-purpose source code (C++), and document markup (LaTeX). ISML supports a range of mathematical models such as those described with ODEs and PDEs as well as ABM-based models that utilize descriptions of conditional behavior.

ISML models a system as an aggregate set of modules corresponding to entities each with a state and corresponding implementation. The implementation details how the states change in reaction to specific events and to the progression of time and, like CellML and SBML, can be modeled mathematically. Graph-like edges linking input/output nodes of modules, termed ports, are used for signaling and communication between modules. These edges enable the communication of physical quantities representing different values of a module's

internal state. By structuring biological models in this way, models can be constructed from components as graphs and hierarchies and represented as independent modules. Modules in ISML are conceptualized as capsules (the authors term this “capsulation”). Capsulation is where multiple capsules are grouped and linked together, packaging them into a larger capsule module. Capsules, like other modules, also possess input and output ports to allow aggregates to be composed of other aggregates. To create logical connections between capsules ports with the encapsulated internal modules, a special case of edge, termed a “forwarding edge,” links capsule ports with internal module ports. Capsulation can lead to the composition of hierarchical representations of models.

A number of similarities exist between ISML and CellML. For example, the concepts of ISML modules and physical quantities map to CellML’s components and variables respectively. However, these parallels are not exhaustive. Although CellML connections between public and private model interfaces are structurally similar to ISML’s edges and ports, respectively, in CellML connections link entity variables that are semantically equivalent but do not model any directionality. ISML edges have explicit direction from input ports to output ports. Additionally, ISML edges can have operational types attached to them by labeling each edge with a verb or verb phrase describing a functional relationship. ISML modules also have a defined type such as functional unit, container, capsule, or template. The definitions of each ISML module type are not discussed in this paper but are described in full in a series of papers by the original authors.^{60–62}

Discussion

Each of the markup languages reviewed approaches modeling different aspects of biology in a generalized fashion. However, we do not believe that such a generic approach to modeling is appropriate when considering more complex, wide-ranging phenomenon and multiscale behaviors considered in the current cancer modeling literature. Typically, these state-of-the-art markup languages take a declarative mathematical approach to modeling, where the biological simulations are derived by mathematical formulae being fed into numerical solvers. They all

use MathML for functional and behavioral description, and, while MathML is a mature markup language, it does not provide any constructs for describing logic and control flow or complex data structures. Models based on markup using MathML are typically simulated through solving ODEs and differential algebraic equations. Control flow constructs and domain-specific components will give a biological modeling markup language more expressive power, especially where models are developed using an *in silico* methodology rather than a purely declarative mathematical approach.⁶³ For example, algorithmic and cellular automata-based cancer models cannot be expressed in any of the currently available markup languages, let alone any hybrid top-down–bottom-up composite models.

The generic application target of these markup languages is also a barrier to their adoption and usage for cancer modeling. SBML is a specialized language that describes molecular components and their relationships with each other. CellML expresses models as declarative mathematics that is processed by numerical solvers, mainly to model cell function, and the domain concepts in CellML are decoupled from the language as metadata annotations. FieldML adds the spatial description element to compliment CellML models, but it is however limited to continuous models of behaviour, being unable to represent the discrete ABM approaches. ISML is similar to CellML in its application to a wide range of biology, and also demonstrates multiscale application, but again in a very generic fashion.

SBML Level 3 supports modular linking of models through hierarchical model composition. A model definition might contain several submodel instances each as part of a composite model. “Ports” are used to act as interfaces to internal elements of a SBML model; however, these are optional. In practice, submodels declared with model definitions embedded within the same document allow direct access to other submodel internal details. Although the use of ports was introduced to define abstract interfaces, they do not enforce any sort of encapsulation ultimately leading to content coupling. CellML version 1.1 is designed to be modular in that encapsulated models can be linked together through public and private interfaces. This allows multiple models whose variables might refer to the same entity to be logically linked.

This component-based approach allows reuse of whole models or parts of models described with CellML markup. This modular coupling has been demonstrated in a number of published models; however, it is not without its problems. Variables can be made directly accessible by declaring them to be so through public interfaces leading to content coupling. Both SBML and CellML encapsulate models and internal components to a certain degree, but their approaches look to relatively basic solutions to ensure backwards compatibility for existing models. What neither language takes into account is that by simply allowing direct connectivity of data between modules, any notion of cohesion is not accounted for. Models typically simulate multiple processes where biological concepts may be spread over different parts of the code or multiple concepts represented in one portion of code. A smarter approach to grouping concerns is needed to achieve true modularity and encapsulation.

To address the specific domain of cancer modeling, we are actively involved with the development of a markup language, TumorML, to describe computational models within the TUMOR project. The motivation for such a markup language is 2-fold: (1) to describe the implementation of these cancer models in an abstract manner that is not tied to any particular programming notation and (2) to be able to couple our models⁶⁴ to address cases such as the GBM scenario described earlier. The challenges posed in developing TumorML include formalizing cancer terminology, linking biological entities with computational and mathematical elements of models, and incorporating features to allow for curating models in online repositories.

Initially we have developed a vocabulary that includes metadata for curation (reusing the Dublin Core Metadata Element Set (<http://www.dublincore.org>) combined with our own cancer-specific metadata) and for describing the public interfaces with existing models that have been developed and published as source code and executable files, the groundwork of which has been described by Johnson et al.⁶⁵ This will allow us to investigate how to couple models of different scales together through their exposed parametric inputs and outputs; an initial “black box” approach to computational model execution and coupling. Portions of the Job Description Markup

Language⁶⁶ are reused to facilitate the specification of the underlying computational requirements for executing computational models. Parametric interfaces are described as named variables with unit and quantity metadata annotations. These computational interfaces could then be mapped to biological terminology ultimately providing a way to more easily validate the cancer biology through correct semantic matching, but also to provide a means to enforce type and units checking where heterogeneities in model descriptions exist.

Domain-specific markup, such as TumorML, could be used to assist in the coupling models of different scales that may have very distinct concerns. If we consider our earlier GBM scenario, our first model is solely concerned with differentiating between real and pseudo progression and regression of a tumor after radiotherapy. This might be calculated by analyzing a patient’s MRI scans at the macroscopic scale. Our second model concerns simulation of predicted tumor growth after a combination of therapies, perhaps looking at cell-cell interactions in tissue at the microscopic scale based on a wide range of patient data in combination with initial MRI scans. Treating model implementations as black boxes, exposed only through a declared interface, in combination with metadata relating to how to run a model, may allow compound models to be constructed via markup.

The cancer modeling community is adopting TumorML for publishing existing models, beginning with efforts in TUMOR and related projects, and we are working with modelers to develop the next level of more detailed abstractions of the inner workings of models, such as work on embedding functional domain-specific code (for vascular tumor growth) into TumorML documents.⁶⁷ Significant effort might be required to port existing models to TumorML, so by providing multiple levels of abstractive notation in our markup we can wrap existing models in early versions of TumorML as well as develop new models with an evolving markup specification. Experimental and clinically oriented vocabularies or ontologies, such as the Simulation Experiment Description Markup Language (SED-ML)⁶⁸ and the CancerGrid model ontology,⁶⁹ could also be integrated to assist in management of clinical trials of TumorML models.



Conclusion

Predictive in silico oncology is an evolving field of study, and, with an increasing number of models being developed by a growing research community, standards need to be adopted to facilitate model sharing and interoperability. Modeling cancer is approached from two distinct angles: bottom-up, from a finer-level molecular modeling detail to simulate higher-level observed behaviors, and top-down, where models are actually based on macroscopic observed behavior. The two types of techniques can be combined to possibly increase the accuracy of clinically relevant models such as those exploited by the VPH projects (ACGT, ContraCancrum) and CViT. An ability to couple models of different scales and approaches is needed, as illustrated by the GBM scenario described in this paper, where the combination of models to differentiate between real progression and pseudoprogression in response to treatment and the simulation of tumor growth with and without treatment is just one possible example of how one might increase the accuracy of using such computational cancer models in predictive oncology. The currently available markup languages reviewed each have their own merits; however, each also has its own pitfalls when applied to the cancer-modeling domain. SBML is specific to the domain of systems biology, FieldML, to physiological structures, while CellML and ISML are too generalized and lack the domain-specific descriptive power required in cancer modeling. TumorML is being developed to address the need for a standardized domain-specific multiscale cancer markup language, where the existing state-of-the-art alternatives fall short. The combination of existing cancer ontologies with existing markup vocabularies will lead to the facilitation of model coupling, ultimately leading to the possibility of better, more accurate in silico models that move one step closer to clinical translation and use in predictive oncology.

Acknowledgements

The authors would like to thank the participating institutions of the project consortium and its members.

Author Contributions

Wrote the first draft of the manuscript: DJ. Contributed to the writing of the manuscript: All authors. Agree with manuscript results and conclusions: All authors. Jointly developed the structure and arguments for the

paper: DJ, SM. Made critical revisions and approved final version: DJ. All authors reviewed and approved of the final manuscript.

Funding

This work was supported in part by the European Commission under the Transatlantic Tumor Model Repositories project (Contract # FP7-ICT-2009.5.4-247754).

Competing Interests

Author(s) disclose no potential conflicts of interest.

Disclosures and Ethics

As a requirement of publication the authors have provided signed confirmation of their compliance with ethical and legal obligations including but not limited to compliance with ICMJE authorship and competing interests guidelines, that the article is neither under consideration for publication nor published elsewhere, of their compliance with legal and ethical guidelines concerning human and animal research participants (if applicable), and that permission has been obtained for reproduction of any copyrighted material. This article was subject to blind, independent, expert peer review. The reviewers reported no competing interests.

References

1. Anderson AR, Quaranta V. Integrative mathematical oncology. *Nat Rev Cancer*. 2008;8(3):227–34.
2. Deisboeck TS, Zhang L, Yoon J, Costa J. In silico cancer modeling: is it ready for prime time? *Nat Clin Pract Oncol*. 2009;6(1):34–42.
3. Ventura AC, Jackson TL, Merajver SD. On the role of cell signaling models in cancer research. *Cancer Res*. 2009;69(2):400–2.
4. Enderling H, Chaplain MA, Anderson AR, Vaidya JS. A mathematical model of breast cancer development, local treatment and recurrence. *J Theor Biol*. 2007;246(2):245–59.
5. Powathil G, Kohandel M, Sivaloganathan S, Oza A, Milosevic M. Mathematical modeling of brain tumors: effects of radiotherapy and chemotherapy. *Phys Med Biol*. 2007;52(11):3291–306.
6. Marias K, Sakkalis V, Roniotis A, et al. Clinically oriented translational cancer multilevel modeling: The ContraCancrum project. In: Dössel O, Schlegel WC, editors. *IFMBE Proceedings 25(4), World Congress on Medical Physics and Biomedical Engineering, Sep 7–12, 2009, Munich, Germany*. Berlin, Germany: Springer-Verlag; 2010:2124–7.
7. Dionysiou D, Stamatakis G, Gintides D, Uzunoglu N, Kyriaki K. Critical parameters determining standard radiotherapy treatment outcome for glioblastoma multiforme: a computer simulation. *Open Biomed Eng J*. 2008;2:43–51.
8. Stamatakis G, Antipas V, Uzunoglu N. A spatiotemporal, patient individualized simulation model of solid tumor response to chemotherapy in vivo: the paradigm of glioblastoma multiforme treated by temozolomide. *IEEE Trans Biomed Eng*. 2006;53(8):1467–77.
9. Rockne R, Alvord EC, Szeto M, Gu S, Chakraborty G, Swanson KR. Modeling diffusely invading brain tumors an individualized approach to quantifying glioma evolution and response to therapy. In: *Selected Topics in Cancer Modeling*. Boston, MA: Birkhäuser; 2008:1–15.

10. Sottoriva A, Verhoe JJ, Borovski T, et al. Cancer stem cell tumor model reveals invasive morphology and increased phenotypical heterogeneity. *Cancer Res.* 2010;70(1):46–56.
11. Frieboes HB, Zheng X, Sun CH, Tromberg B, Gatenby R, Cristini V. An integrated computational/experimental model of tumor invasion. *Cancer Res.* 2006;66(3):1597–604.
12. Castorina P, Carcò D, Guiot C, Deisboeck TS. Tumor growth instability and its implications for chemotherapy. *Cancer Res.* 2009;69:8507–15.
13. Sakkalis V, Roniotis A, Farmaki C, Karatzanis I, Marias K. Evaluation framework for the multilevel macroscopic models of solid tumor growth in the glioma case. *Conf Proc IEEE Eng Med Biol Soc.* 2010;2010: 6809–12.
14. Roniotis A, Marias K, Sakkalis V, Zervakis M. Diffusive modeling of glioma evolution: a review. *J Biomed Sci Eng.* 2010;3(5):501–8.
15. Dionysiou D, Stamatakis G, Uzunoglu N, Nikita KS, Marioli A. A four-dimensional simulation model of tumour response to radiotherapy in vivo: parametric validation considering radiosensitivity, genetic profile and fractionation. *J Theor Biol.* 2004;230(1):1–20.
16. Dionysiou D, Stamatakis G. Applying a 4D multiscale in vivo tumor growth model: investigating the effect of tumour oxygenation on radiation therapy response. *Cancer Inform.* 2007;2:113–21.
17. Ubezio P, Cameron D. Cell killing and resistance in pre-operative breast cancer chemotherapy. *BMC Cancer.* 2008;8:201.
18. Titz B, Jeraj R. An imaging-based tumour growth and treatment response model: investigating the effect of tumour oxygenation on radiation therapy response. *Phys Med Biol.* 2008;53(17):4471–88.
19. Anderson AR, Weaver AM, Cummings PT, Quaranta V. Tumor morphology and phenotypic evolution driven by selective pressure from the microenvironment. *Cell.* 2006;127(5):905–15.
20. Duechting W, Ulmer W, Lehigh R, Ginsberg T, Dedeleit E. Computer simulation and modelling of tumor spheroid growth and their relevance for optimization of fractionated radiotherapy. *Strahlenther Onkol.* 1992;168(6):354–60.
21. Norris V, Cabin A, Zermiline A. Hypercomplexity. *Acta Biotheor.* 2005; 53(4):313–30.
22. Deisboeck TS. Personalizing medicine: a systems biology perspective. *Mol Syst Biol.* 2009;5:249.
23. Lowengrub JS, Frieboes HB, Jin F, et al. Nonlinear modelling of cancer: bridging the gap between cells and tumours. *Nonlinearity.* 2010; 23(1):R1–9.
24. Sanga S, Frieboes HB, Zheng X, Gatenby R, Bearer EL, Cristini V. Predictive oncology: a review of multidisciplinary, multiscale in silico modeling linking phenotype, morphology and growth. *Neuroimage.* 2007;37(Suppl 1):S120–34.
25. Wang Z, Deisboeck TS. Computational modeling of brain tumors: discrete, continuum or hybrid? *Sci Model Simul.* 2008;15:381–93.
26. Noble D. The rise of computational biology. *Nat Rev Mol Cell Biol.* 2002;3(6):459–63.
27. Wang Z, Zhang L, Sagotsky J, Deisboeck TS. Simulating non-small cell lung cancer with a multiscale agent-based model. *Theor Biol Med Model.* 2007;4:50.
28. Wang Z, Birch CM, Deisboeck TS. Cross-scale sensitivity analysis of a non-small cell lung cancer model: Linking molecular signaling properties to cellular behavior. *Biosystems.* 2008;92(3):249–58.
29. Wang Z, Birch CM, Sagotsky J, Deisboeck TS. Cross-scale, cross-pathway evaluation using an agent-based non-small cell lung cancer model. *Bioinformatics.* 2009;25(18):2389–96.
30. Wang Z, Bordas V, Sagotsky J, Deisboeck TS. Identifying therapeutic targets in a combined EGFR-TGF β R signalling cascade using a multiscale agent-based cancer model. *Math Med Biol.* 2012;1:95–108.
31. Gerlee P, Anderson A. An evolutionary hybrid cellular automaton model of solid tumour growth. *J Theor Biol.* 2007;246(4):583–603.
32. Gerlee P, Anderson A. A hybrid cellular automaton model of clonal evolution in cancer: the emergence of the glycolytic phenotype. *J Theor Biol.* 2008;250(4):705–22.
33. Gerlee P, Anderson A. Evolution of cell motility in an individual-based model of tumour growth. *J Theor Biol.* 2009;259(1):67–83.
34. Ramis-Conde I, Drasdo D, Anderson AR, Chaplain MA. Modeling the influence of the E-Cadherin- β -Catenin pathway in cancer cell invasion: a multiscale approach. *Biophys J.* 2008;95(1):155–65.
35. Ramis-Conde I, Chaplain MA, Anderson AR, Drasdo D. Multi-scale modelling of cancer cell intravasation: the role of cadherins in metastasis. *Phys Biol.* 2009;6(1):016008.
36. Stamatakis GS, Dionysiou D, Graf N, et al. The “Oncosimulator”: a multilevel, clinically oriented simulation system of tumor growth and organism response to therapeutic schemes. Towards the clinical evaluation of in silico oncology. *Conf Proc IEEE Eng Med Biol Soc.* 2007;2007: 6629–32.
37. Graf N, Hoppe A, Georgiadi E, et al. ‘In silico’ oncology for clinical decision making in the context of nephroblastoma. *Klin Padiatr.* 2009; 221(3):141–9.
38. Clapworthy GJ, Kohl P, Gregerson H, et al. Digital human modelling: a global vision and a European perspective. In: *Digital Human Modeling, Lecture Notes in Computer Science.* 2007;4561:549–58.
39. Martin L, Anguita A, Graf N, et al. ACGT: advancing clinico-genomic trials on cancer—four years of experience. *Stud Health Technol Inform.* 2011;169: 734–8.
40. Deisboeck TS, Zhang L, Martin S. Advancing cancer systems biology: introducing the Center for the Development of a Virtual Tumor, CViT. *Cancer Inform.* 2007;5:1–8.
41. Sakkalis V, Marias K, Stamatakis G, et al. The TUMOR project: integrating cancer model repositories for supporting predictive oncology. In: *Abstract Booklet for VPH2012 Integrative Approaches to Computational Biomedicine.* September 18–20, 2012; London, UK.
42. Louis D, Ohgaki H, Wiestler OD, et al. The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathol.* 2007; 114(2):97–109.
43. Wen PY, Kesari S. Malignant gliomas in adults. *N Engl J Med.* 2008; 359(5):492–507.
44. Warmuth-Metz M. Neuroradiologie bei gliomen und metastasen. *Der Onkologe.* 2011;17(1):18–30.
45. Scott JN, Brasher PM, Sevic RJ, Rewcastle NB, Forsyth PA. How often are nonenhancing supratentorial gliomas malignant? A population study. *Neurology.* 2002;59(6):947–9.
46. Rees J. Glioma therapy. *ACNR.* 2002;2(2):11.
47. Stupp R, Mason WP, van den Bent MJ, et al. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N Engl J Med.* 2005; 352(10):987–96.
48. Krex D, Klink B, Hartmann C, von Deimling A, et al. Long-term survival with glioblastoma multiforme. *Brain.* 2007;130(Pt 10):2596–606.
49. Hegi ME, Diserens AC, Gorlia T, et al. MGMT gene silencing and benefit from temozolomide in glioblastoma. *N Engl J Med.* 2005;352(10):997–1003.
50. Wen PY, Macdonald DR, Reardon DA, et al. Updated response assessment criteria for high-grade gliomas: response assessment in neuro-oncology working group. *J Clin Oncol.* 2010;28(11):1963–72.
51. Zepp J, Graf N, Skounakis E, et al. Tumor segmentation: The impact of standardized signal intensity histograms in glioblastoma. In: *Proceedings of the 4th International Advanced Research Workshop on In Silico Oncology and Cancer Investigation.* Athens, Greece; Sep 8–9, 2010:98–9.
52. Bray T, Paoli J, Sperberg-McQueen C, Maler E, Yergeau F, editors. *Extensible Markup Language (XML) 1.0*, 5th ed. *W3C Recommendation 26 Nov 2008.* <http://www.w3.org/TR/2008/REC-xml-20081126/>. Accessed April 18, 2013.
53. Wiggins R. XML: An interview with Peter Flynn. *IEEE Comput.* 2000;33(4):113–6.
54. Hucka M, Finney A, Bornstein BJ, et al. Evolving a lingua franca and associated software infrastructure for computational systems biology: the Systems Biology Markup Language (SBML) project. *IET Syst Biol.* 2004;1(1):41–53.
55. Hucka M, Bergmann F, Hoops S, Keating SM, Sahle S, Wilkinson DJ. The Systems Biology Markup Language (SBML): language specification for level 3 version 1 Core (release 1 candidate). *Nature Precedings.* 2010. <http://dx.doi.org/10.1038/npre.2010.4123.1>. Accessed Jan 8, 2013.
56. Carlisle D, Ion P, Miner R, et al. Mathematical Markup Language (MathML) version 3.0. *W3C Recommendation 21.* 2010.



57. Cuellar AA, Lloyd CM, Nielsen PF, Bullivant DP, Nickerson DP, Hunter PJ. An overview of CellML 1.1, a biological model description language. *SIMUL.* 2003;79(12):740–7.
58. Lloyd CM, Halstead MD, Nielsen PF. CellML: its future, present and past. *Prog Biophys Mol Biol.* 2004;85(2–3):433–50.
59. Christie GR, Nielsen PM, Blackett SA, Bradley CP, Hunter PJ. FieldML: concepts and implementation. *Philos Transact A Math Phys Eng Sci.* 2009; 367(1895):1869–84.
60. Asai Y, Suzuki Y, Kido Y, et al. Specifications of insilicoML 1.0: a multilevel biophysical model description language. *J Physiol Sci.* 2008;58(7):447–58.
61. Suzuki Y, Asai Y, Kawazu T, et al. A platform for in silico modeling of physiological systems II. CellML compatibility and other extended capabilities. *Conf Proc IEEE Eng Med Biol Soc.* 2008;2008:573–6.
62. Suzuki Y, Asai Y, Oka H, et al. A platform for in silico modeling of physiological systems III. *Conf Proc IEEE Eng Med Biol Soc.* 2009;2009: 2803–6.
63. Johnson D, McKeever S. The case for using markup for biomechanical modelling. In: *Proc Internat Soc Biomech XXIII World Congr.* 2011.
64. Johnson D, Cooper J, McKeever S. Markup languages for in silico oncology, In: *Proceedings of the 4th International Advanced Research Workshop on In Silico Oncology and Cancer Investigation.* Athens, Greece, Sep 8–9, 2010:108–10.
65. Johnson D, Cooper J, McKeever S. TumorML: Concept and requirements of an in silico cancer modelling markup language. *Conf Proc IEEE Eng Med Biol Soc.* 2011;2011:441–4.
66. Anjomshoa A, Brisard F, Drescher M, et al. Job Submission Description Language (JSDL) Specification, Version 1.0. Technical Report. *Global Grid Forum.* 2005.
67. Johnson D, Connor AJ, McKeever S. Modular markup for simulating vascular tumour growth. In: *Proceedings of the 4th International Advanced Research Workshop on In Silico Oncology and Cancer Investigation.* Athens, Greece; Sep 8–9, 2010:53–6.
68. Köhn D, Novère N. SED-ML—an XML format for the implementation of the MIASE guidelines. In: *Computational Methods in Systems Biology, Lecture Notes in Computer Science.* 2008;5307:176–90.
69. Calinescu R, Harris S, Gibbons J, Davies J. Cross-trial query system for cancer clinical trials, In: Sobh T, editor. *Innovations and Advanced Techniques in Computer and Information Sciences and Engineering.* Dordrecht, The Netherlands: Springer Netherlands; 2007:385–90.